AD-A216 748

ANALYSIS OF FIRST-TERM ATTRITION OF NON-PRIOR SERVICE HIGH-QUALITY U.S. ARMY MALE RECRUITS

by

HOA GENERAZIO

B.S., United States Military Academy
(1977)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN OPERATIONS RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

December 1989

Signature of Author _____
Operations Research Center
December 1989

Certified by _____
Arnold I. Barnett
Professor of Operations Research and Management
Thesis Advisor

Accepted by _____
Amedeo R. Odoni
Professor of Aeronautics and Astronautics and of Civil Engineering
Co-director, Operations Research Center

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	13 December 1989	Final

4. TITLE AND SUBTITLE
Analysis of First-Term Attrition of Non-Prior Service High-Quality U.S. Army Male Recruits

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ Hoa Generazio

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
HQDA, MILPERCEN (DAPC-OPB-D)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

An analysis is performed to estimate an individual's probability of attrition in terms of certain of his characteristics at time of enlistment. The main technique is logistic regression modelling, which is applied to data pertaining to the high-quality male population of the U.S. Army FY84 NPS accession cohort (high-school graduates who scored 50% or higher on the Armed Forces Qualification Test (AFQT). The results showed four significant characteristics are: age, level of education, aptitude test score, and entry status (with or without a waiver). Age and entry status were positively correlated with the rate of attrition. Conversely, education and aptitude test score were negatively correlated. The older the recruit, the more likely the person is to drop out. The recruit also is in a higher risk category for attrition if he or she entered the Army with a waiver. The better educated the recruit is, the less likely the person will drop out. The higher the aptitude test score, the more likely that the recruit will remain for the entire obligated tour.

14. SUBJECT TERMS	15. NUMBER OF PAGES
First Term Attrition; Non-Prior Service; FY84 Accession Cohort Significant Characteristics; Logistic Regression; Maximum Likelihood Estimation	50
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASS	UNCLASS	UNCLASS	UL

ANALYSIS OF FIRST TERM ATTRITION OF NON-PRIOR SERVICE HIGH-QUALITY U.S. ARMY MALE RECRUITS

by

HOA GENERAZIO

Submitted to the Operations Research Center on December 13, 1989
in partial fulfillment of the requirements for
the degree of Master of Science in Operations Research

ABSTRACT

The United States Army enlists an average of 120,000 recruits (90% male) each year to sustain its military force. Unfortunately, an average of 30% of each non-prior service (NPS) accession cohort (year group) will , voluntarily or involuntarily, depart the Army prior to completing its initial tour length obligation. For military readiness and fiscal reasons, reducing attrition is a high concern of Army leaders.

An analysis is performed to estimate an individual's probability of attrition in terms of certain of his characteristics at time of enlistment. The main analytic technique is logistic regression modelling, which is applied to data pertaining to the high quality male population of the U.S. Army FY84 NPS accession cohort (high school graduates who scored 50% or higher on the Armed Forces Qualification Test (AFQT)).

The results of the analysis showed four significant characteristics are: age, level of education, aptitude test score, and entry status (with or without a waiver).
Age and entry status were positively correlated with the rate of attrition. Conversely, education and aptitude test score were negatively correlated. The older the recruit, the more likely the person is to drop out. The recruit also is in a higher risk category for attrition if he or she entered the Army with a waiver. The better educated the recruit is, the less likely the person will drop out. The higher the aptitude test score, the more likely that the recruit will remain for the entire obligated tour.

Thesis Supervisor: Professor Arnold I. Barnett
Title: Professor of Operations Research and Management

ACKNOWLEDGEMENTS

I want to thank my wife, Susan, and my three children, Christopher, Andrew and Alaina, for their love and support.

I want to thank my academic advisor and thesis supervisor, Arnie Barnett. His patient, support and constructive comments guided my work from the beginning to end.

I want to thank The Department of Mathematics, United States Military Academy for giving me the opportunity to pursue a masters' degree in Operations Research.

I want to thank Ilze S. Levis for her skillful and critical reviews, of my numerous drafts, for technical style, logical organization and overall quality.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 BACKGROUND

The United States Army enlists an average of 120,000 non-prior service recruits (90% male) each year to sustain its military force. These enlistees receive training to qualify them for a military occupational specialty (MOS). Depending upon the particular MOS, the training can last from four months to a year. Unfortunately, an average of 30% of each non-prior service (NPS) accession cohort (year group) will , voluntarily or involuntarily, depart the Army prior to completing its initial tour length obligation.

First-term attrition is very costly to the U.S. Army, since force size and readiness is affected, while recruiting and training costs are not fully recovered. Army leaders and manpower planners are concerned about the high attrition rates, and numerous studies have been done to find explanations (Buddin, 1984; Toomepuu, 1986; Horne, 1986).

One area of research has concentrated on determining if the characteristics of a recruit, at time of enlistment, are significant factors in determining if the person is a high or a low risk for attrition. Studies have shown that those individuals categorized by the U.S. Army as high-quality recruits (high school graduates who scored 50% or higher

on the Armed Forces Qualification Test (AFQT)[1]) have attrition rates at about half those of the other recruits in that cohort (Fernandez, 1985; Armor, 1982; Sinaiko, 1981). An average of 55% per accession cohort are classified as high quality. Extensive studies have also shown that male recruits have markedly different attrition patterns and lower attrition rates than their female counterparts (Quester, 1986; Grissmer, 1985).

1.2 APPROACH

This study differs from previous ones in that it examines only the non-prior service high quality male population of an accession cohort. It attempts to determine, within this " desirable " recruiting population, if any characteristics of a recruit can be quantifiable factors in categorizing the person as a high or low risk for attrition.

The reasons for analysis of this particular group are twofold. First-term attrition among high-quality recruits is especially costly since these recruits have higher associated recruiting and training costs than those of other recruits. The male population was selected since it generally makes up between 85% to 90% of the total cohort population.

This study used a logistic regression model and maximum likelihood estimation for estimating parameter values. First, a large sample population was randomly chosen

---

[1] The score is a composite of a subset of the individual Armed Services Vocational Aptitude Battery (ASVAB) component scores, reflecting language and arithmetic skills, and is used as a measure of general aptitude, trainability and productivity.

from the entire accession cohort population. As a baseline, a Chi-square test was conducted to rule out the possibility that variation in the attrition rates among the different recruiting groups were caused by merely random noise. A multivariate logistic regression model was then developed to attempt to quantify the effects of the variables considered on attrition levels. Maximum likelihood estimation was used to obtain the best estimation of model parameters. Model validation was accomplished with the 5% hold-out sample. Once the model was validated, it was used to estimate the probability of attrition for individuals with specific characteristics. The results can be used by Army leaders and manpower planners to determine criteria for future enlistment and to minimize attrition.

1.3 ORGANIZATION OF THESIS

The next three sections are organized as follows. Section II describes the analysis methods and data base used in the study. Section III shows the results about the attrition patterns for high-quality, male recruits of FY84 cohort. Section IV reports the conclusions.

Chapter 2

ANALYSIS METHODS AND DATA BASE

Attrition was modeled as a function of certain individual traits. A Chi-square test was conducted to assure that the variation of attrition rates exhibited by the different groups were not attributed solely to randomness, and that an attempt to explain the phenomenon via a mathematical model was worthwhile and appropriate.

A multivariate logistic regression model was developed to assess the importance of the variables hypothesized to influence attrition. Maximum likelihood estimation was used to estimate the parameter values of the model, and allowed the model to be used to estimate the probability of attrition for an individual with specific characteristics.

2.1 SELECTION OF VARIABLES

Four characteristic traits were chosen to use in the model. These particular traits -- age, education level, aptitude test score category, and entry status -- all have one common and important attribute; the Army already has recognized the importance of their effects on force readiness and has set entrance criteria and limitations for each trait. Thus, if the results of this study show that these traits also affect the attrition rate, then existing regulations can be changed or the focus of recruiting can be directed towards a particular segment of the population, within the set guidelines and directives. Variables, deemed discriminatory, such as " race " and " number of

dependents ", or those deemed not changeable by policy, such as " home state ", were not studied, since a finding would be of interest, but of minimal value.

Table 1 provides the definitions of the variables used in the multivariate analysis.

TABLE 1

ASSIGNMENT AND DEFINITIONS OF VARIABLES

Variable	Definition
Age	
X11	Age group 17 - 21
X12	Age group 22- 25
X13	Age group 26 - 29
X14	Age group >= 30
Education Level	
X21	High School graduate
X22	High School graduate with <= 2 yrs college
X23	High School graduate with > 2 yrs college
Aptitude Category	
X31	Cat IIIA, scored between 50% and 64%
X32	Cat II , scored between 65% and 92%
X33	Cat I , scored above 92%
Entry Status	
X41	Entry without waiver
X42	Entry with waiver (medical, legal ...)

2.2 CHI-SQUARE TEST

Recruits are grouped together by same age, education level, aptitude test score category, and entry status. Suppose there are k cells, and one hypothesizes that the probability of attrition for each cell i, p_i , where i = 1,2....k, is equal to p, where p is the probability of attrition of the entire population. Under this hypothesis, the differences displayed among the observed individual cell attrition rate would be attributed to randomness. The major question is: does $p_i = p$?

The Chi-square test is used on the data to test this hypothesis.

$$\begin{aligned} &\text{Ho}: p_i = p \quad \text{for all } i \\ &H_1: p_i \neq p \end{aligned} \tag{2.1}$$

where: i = 1, ... k (number of cells)

$p \neq 0$

One can obtain a good estimate $\hat{p}$ of p, and let $D(\hat{p}, p_i)$ be a measure of the distance between $\hat{p}$ and p_i . Reject the hypothesis if D is too large; otherwise accept.

Under the observed probability distribution $(\hat{p}_1,\hat{p}_2,...,\hat{p}_k)$, for large n

$$D = \sum_{1}^{k} \frac{n(\hat{p}_k - \hat{p})^2}{\hat{p}} \tag{2.2}$$

has approximately a χ^2_{k-1} distribution, and the test based on D is called the Chi-square test (Breiman, 1973) .

D is compared with a test computed from a χ^2_{k-1} distribution. If D is too large, i.e. greater than the value from the Chi-square table for χ^2_{k-1} at the chosen acceptance region, then H_0 is rejected.

Chi-square tables normally go up to 30 degrees of freedom (k-1 = 30). For k-1 > 30, the acceptance region for D is calculated as:

$$D \leq (k - 1) + z \sqrt{(k - 1)} \qquad (2.3)$$

where z is computed from an N(0,1) table (Breiman, 1973).
The result was used as a baseline to determine whether a logistic regression was likely to show statistically significant predictors of the attrition rate.

2.3 LOGISTIC REGRESSION (LOGIT)

A recruit either separates prior to, or remains in service until, completion of the obligated tour. Let us consider attrition , Y_i , as a binary (two-valued) dependent variable, and attempt to explain it from the independent variables , X_{ij}'s (individual characteristic traits).
The dependent or response variable , Y_i , has a Bernoulli distribution, and is defined as :

Y_i = 1, if individual i did not complete his obligated tour

0, if individual i did complete his obligated tour

and

$$E (Y_i = 1) = prob (Y_i = 1 \mid X_{ij}) \qquad (2.4)$$
$$E (Y_i = 0) = prob (Y_i = 0 \mid X_{ij})$$

The binary observations were available on n recruits, assumed to be independent. The problem was to use the data to develop a good method of analysis for assessing the dependence of the probability of attrition , on X_{ij} , the characteristic traits. The usual linear regression models would be unsatisfactory for two primary reasons:

1. Ordinary least squares linear regression models require that the variance of Y_i be constant, whereas in this case var (Y_i) is a function of the expected Y_i.

2. Linear regression models do not have the constraint $0 \leq p \leq 1$ that this case requires. Linear regression models could thus produce absurd results.

Regression of a logistic response function is more appropriate, since it allows for the variance to depend on the mean p , and it does not allow the estimated probabilities to fall outside the 0 - 1 range. Figure 1 illustrates the S-shaped curve of the logistic response function, which asymptotically approaches zero at one end, and one at the other end.

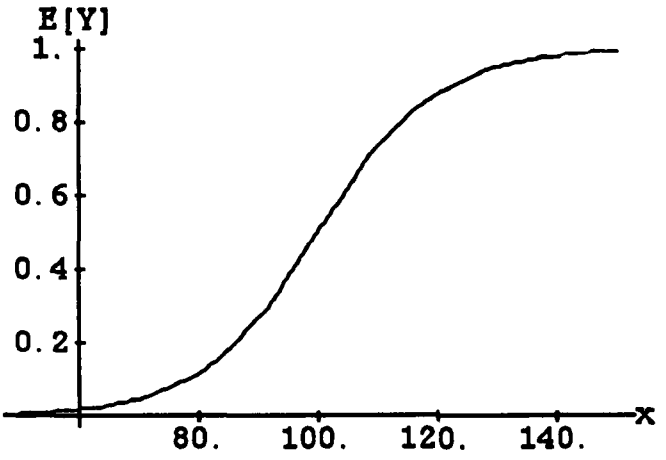

Figure 1 : Logistic Response Function

The model is represented as:

$$E(Y_i = 1) = prob(Y_i = 1 \mid X_{ij}) = \frac{e^{(\beta_o + \sum \beta_j X_{ij})}}{1 + e^{(\beta_o + \sum \beta_j X_{ij})}} \qquad (2.5)$$

where:

X_{ij} = vector of characteristics of a recruit

$prob(Y_i = 1 \mid X_{ij})$ = probability of attrition of a recruit i, with characteristic traits X_{ij}

ß ' s = the parameters of the model.

The observed values of Y_i can be used to fit parameters to this curve and thus give estimates $b_0 \ldots b_j$ of $\beta_0 \ldots \beta_j$, and $\hat{p}$ for $prob(Y_i = 1 \mid X_{ij})$.

One can take a log transformation , called logit, and linearize the statistics:

$$\text{Logit} = \log\left(prob(Y_i = 1 \mid X_{ij}) / (1 - prob(Y_i = 1 \mid X_{ij})\right) = b_0 + \sum b_j X_{ij} \qquad (2.6)$$

Thus, $prob(Y_i = 1 \mid X_{ij})$ is assumed not linear in x; instead, it is the logistic transformation that is assumed linear in x. However, this technique generally gives satisfactory estimates only for grouped data, where the number in each cell is required to be large, and only when each $prob(Y_i = 1 \mid X_{ij})$ is sufficiently far from 0 and 1 so

that the observed number in the cell is approximately normal. In this situation there existed only 8 of 72 cells with small population and/or *prob* ($Y_i = 1 \mid X_{ij}$) at 0 and 1. Nevertheless, I preferred not to transform the data, but to solve the nonlinear equations directly, using maximum likelihood estimation (Morris, 1981).

2.4 MAXIMUM LIKELIHOOD ESTIMATION (MLE)

The major advantage of the maximum likelihood method resides in the asymptotic properties of the estimators. Under broad conditions (Hanushek, 1977), the maximum likelihood estimators are:

(a) consistent

(b) asymptotically efficient, and

(c) asymptotically normal

The principle of maximum likelihood involves choosing b's as estimates of ß's , so that if they were actually ß 's, the given observations would have the highest probability of occurrence.

The maximum likelihood estimation was accomplished with the statistical software package SYSTAT . The procedure used the following log-likelihood function :

$$L(b_n) = \Sigma \, [\, (\text{failure})(\log(\text{estimate}) + (1\text{-failure})(\log(1\text{-estimate}))] \qquad (2.7)$$

where:
bn = the parameters to be estimated
failure = 0 if individual is counted as attrition,
= 1 if individual remained in service

$$\text{estimate} = \frac{e^{(\beta_o + \sum \beta_j X_{ij})}}{1 + e^{(\beta_o + \sum \beta_j X_{ij})}}$$

$\sum$ = sum over all k cells

SYSTAT's MLE is accomplished via minimizing the negative of the log-likelihood function. The estimation is done as follows : First, a model and a loss function are specified. In this case, the model is a logit model and the loss function is expressed as the negative of the log-likelihood function. Starting values for the parameters can be entered or a set of default values can be used. The model is evaluated by using the starting values, the result is called the estimate. The loss statement , in turn, is evaluated using the estimate values. The procedure is repeated for all the cells in the file and the loss is summed over all cells. The summed loss is then minimized via the Quasi-Newton algorithm. This minimization algorithm uses numerical methods of the first and second derivatives of the loss function to seek a minimum. Iterations continue until the tolerance criterion for convergence (tolerance = .00001) is reached.

2.5 DATA BASE

The primary source of personnel information was maintained and provided by the Defense Manpower Data Center (DMDC). The DMDC cohort file for each fiscal year contains the personal, educational, geographical, and professional data, from entry until separation, of all individuals who were identified as gains during that fiscal year. The FY84 cohort file was used because it was the most recent cohort that has data collected for a minimum of 4 years. Recruits can enlist for up to 4 years of service, and we wanted to consider the actual attrition status of those personnel studied.

The entire FY84 cohort NPS high-quality male recruit population was 51,140. The model was calibrated based on data concerning 95% (48,583)of the population. The other 5% (2,557) were kept as a " hold-out " sample for model validation.

Chapter 3

ATTRITION ANALYSIS

3. 1 DEMOGRAPHIC PROFILE - GENERAL

The FY84 accession cohort was composed of 131,933 persons, of which 114,681 (86.9%) were males and 17,252 (13.1%) were females. There were 61,994 (47.0%) high-quality recruits. Male high-quality recruits numbered 51,140 (44.6% of cohort males) and females numbered 10,854 (62.9% of cohort females). Within the high-quality population, 82.5% were males.

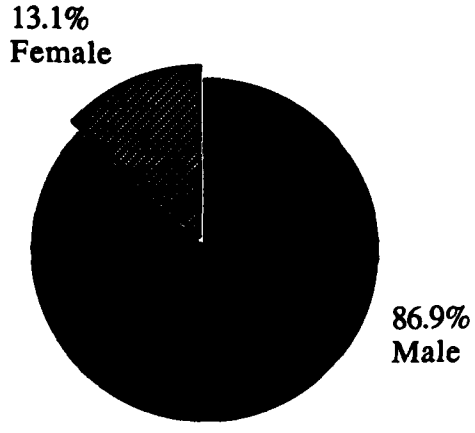

Figure 3.1 : FY84 NPS Accession Cohort
Sex

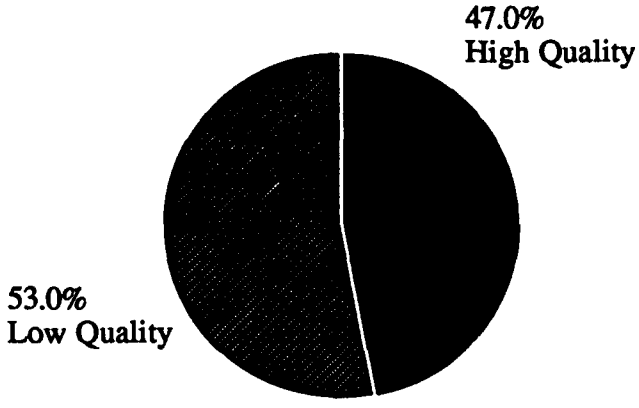

Figure 3.2 : FY84 NPS Accession Cohort
High Quality vs Low Quality

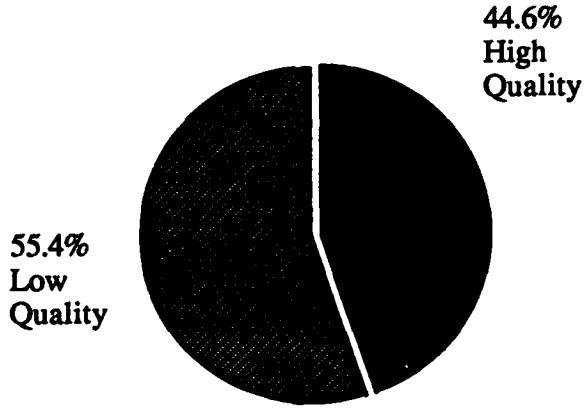

Figure 3.3 : FY84 NPS Accession Cohort
Male - High vs Low Quality

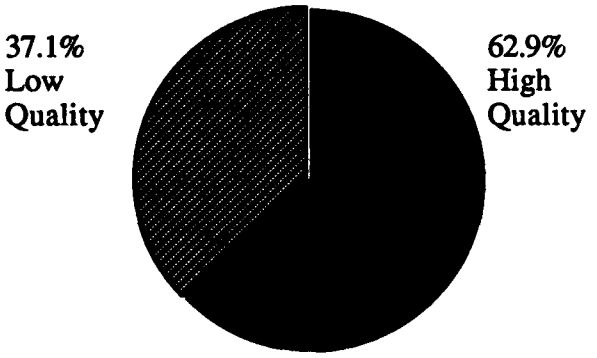

Figure 3.4 : FY84 NPS Accession Cohort
Female - High vs Low Quality

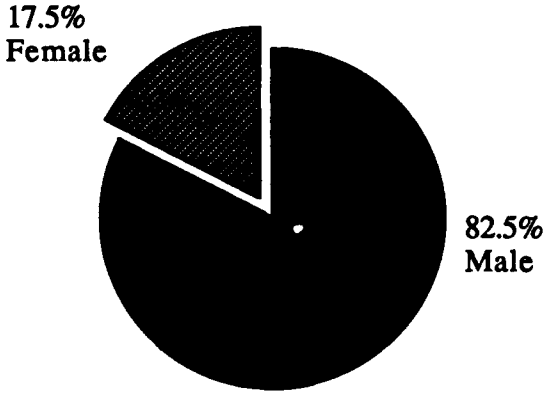

Figure 3.5 : FY84 NPS Accession Cohort
High Quality - Male vs Female

3.2 DEMOGRAPHIC PROFILE - HIGH-QUALITY MALES

There were 11,354 high-quality male recruits who separated from the Army prior to completion of their obligated tours. The attrition rate for this group was 22.2%, as compared to the 30.2% for the overall accession cohort.

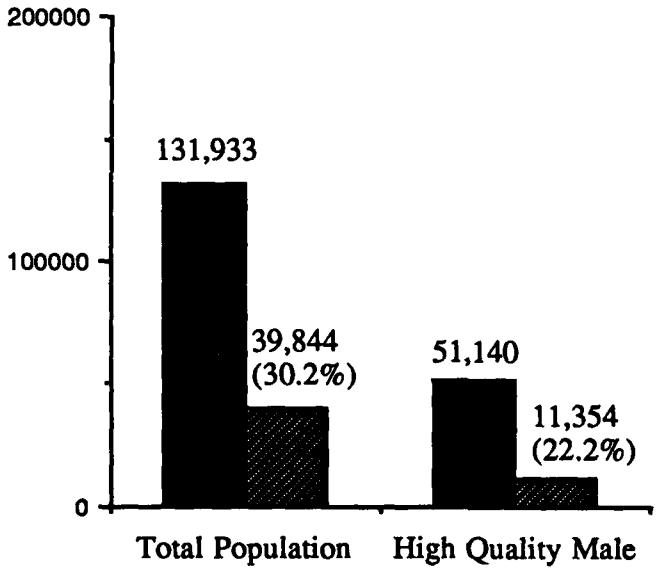

Figure 3.6 : FY84 NPS Accession Cohort Attrition Rates - High Quality Male vs Total Population

3.3 DEMOGRAPHIC PROFILE - HIGH QUALITY MALES - SAMPLE GROUP

3.3.1 AGE

Recruits' ages ranged from 17 to 35, the allowable limits for enlistment. The two ages with the highest enlistment numbers were age 18, with 17347 (35.7%) and age 19, with 8870 (18.3%). The six highest attrition rates belonged to ages 30 - 35. The profiles are shown at Figures 3.7 and 3.8. The recruits were then separated into four major age groups: 17 - 21 (prime - target for recruiting), 22 - 25, 26 - 29, and 30 +. Their attrition rates are 21.8%, 21.9%, 26.8%, and 36.1%, respectively. Figure 3.9 displays the age group populations, and the associated attrition proportions are shown at Figure 10. Generally, the attrition rate increased as age increased. A recruit in an older age group had a higher rate of attrition than a recruit, with similar characteristic traits, of a younger age group.

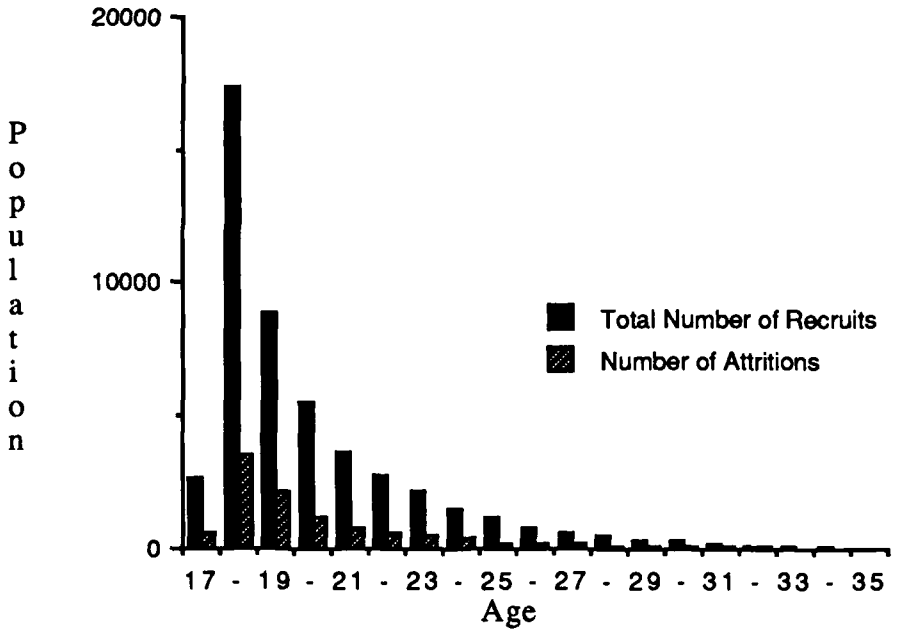

Figure 3.7 : FY84 NPS Accession Cohort - Sample Group - by Age - Total Numbers vs Numbers of Attrition

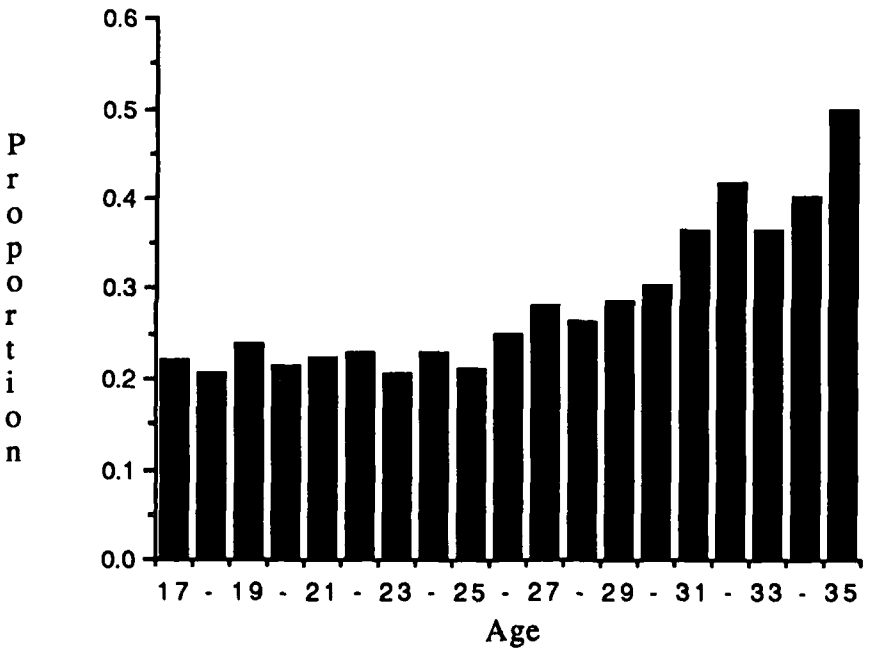

Figure 3.8 : FY84 NPS Sample Group - by Age - Proportion of Attrition

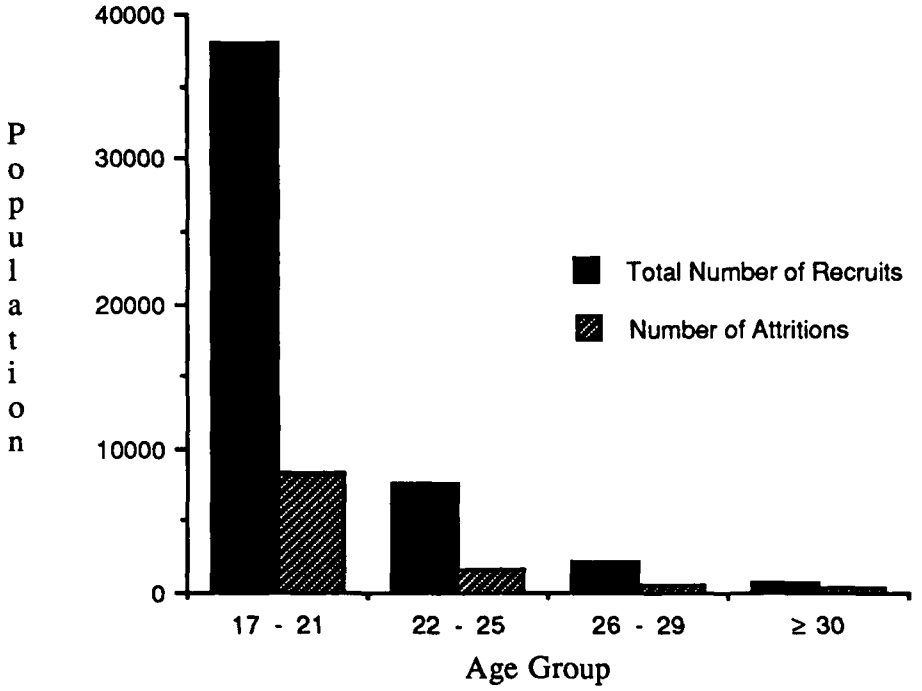

Figure 3.9 : FY84 NPS Accession Cohort - Sample Group - by Age Groups - Total Numbers vs Numbers of Attrition

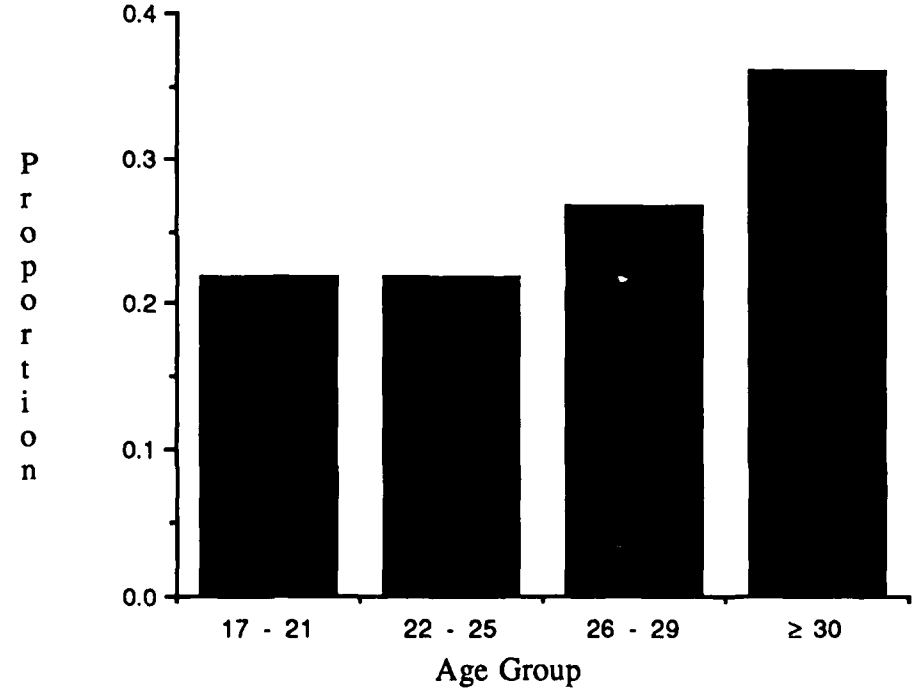

Figure 3.10 : FY84 NPS Accession Cohort - Sample Group - by Age Groups - Proportion of Attrition

3.3.2 EDUCATION LEVEL

High quality recruits' education levels [1] ranged from high school graduates to PhD's. The two education levels with the highest enlistment numbers were high school graduate, with 40,505 (83.4%) and high school graduate with one year of college education, with 2,962 (6.1%). The highest attrition rate belonged to the high school graduate level, at 23.3%. The data is displayed at Figures 3.11 and 3.12. The recruits were combined into three major education level groups: high school graduate, high school graduate with two or fewer years of college education, and high school graduate with more than two years of college education. Their attrition rates are 23.3%, 17.7%, and 15.6%, respectively. The education level groups and their respective attrition rates are graphically displayed by Figure 3.13 and 3.14. Generally, the attrition rate decreased as the education level increased. A recruit in a lower education level group had a higher risk of attrition than a recruit, with similar characteristic traits, of a higher education level group.

One must be careful with this result since age and education level are correlated. As one can reasonably expect, an older person is more likely to have a higher education level and vise versa. But as one is older, one is less likely to remain for the full service tour. Therefore, there may be cancelling effects on-going. A move in one direction for age is associated with the a move in the opposite direction for education level.

---

[1] 6 = High School Graduate (HSG); 7 = HSG + 1 yr college; 8 = HSG + 2 yr college; 9 = HSG + 3 or 4 yrs college (No degree); 10 = College Grad; 11 = Masters; 12 = PhD;

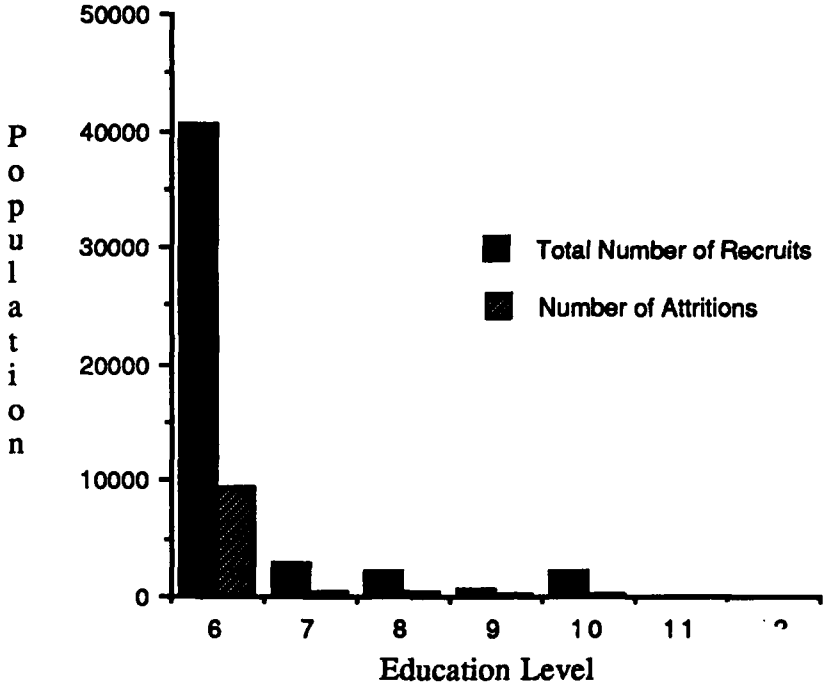

Figure 3.11 : FY84 NPS Accession Cohort - Sample Group - by Education Level - Total Numbers vs Numbers of Attrition

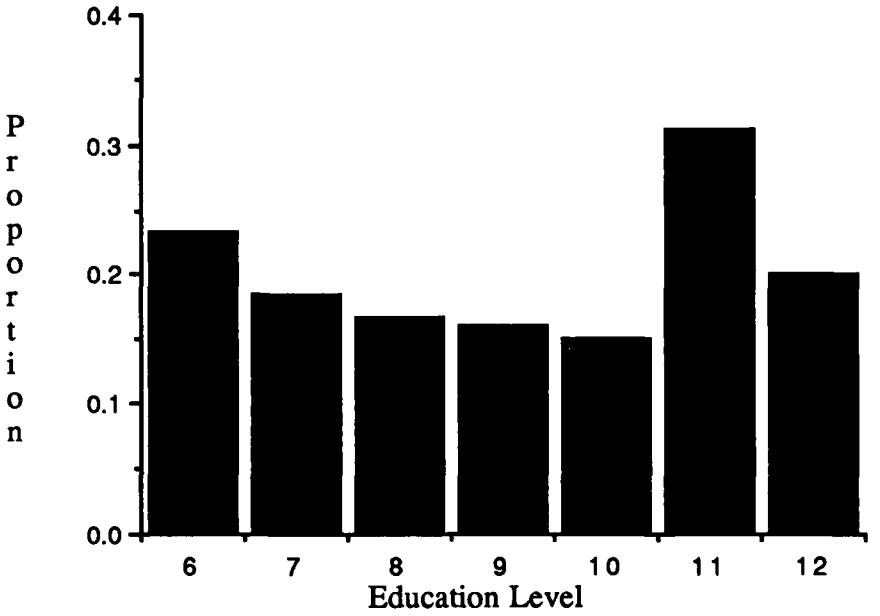

Figure 3.12 : FY84 NPS Accession Cohort - Sample Group - by Education Level - Proportion of Attrition

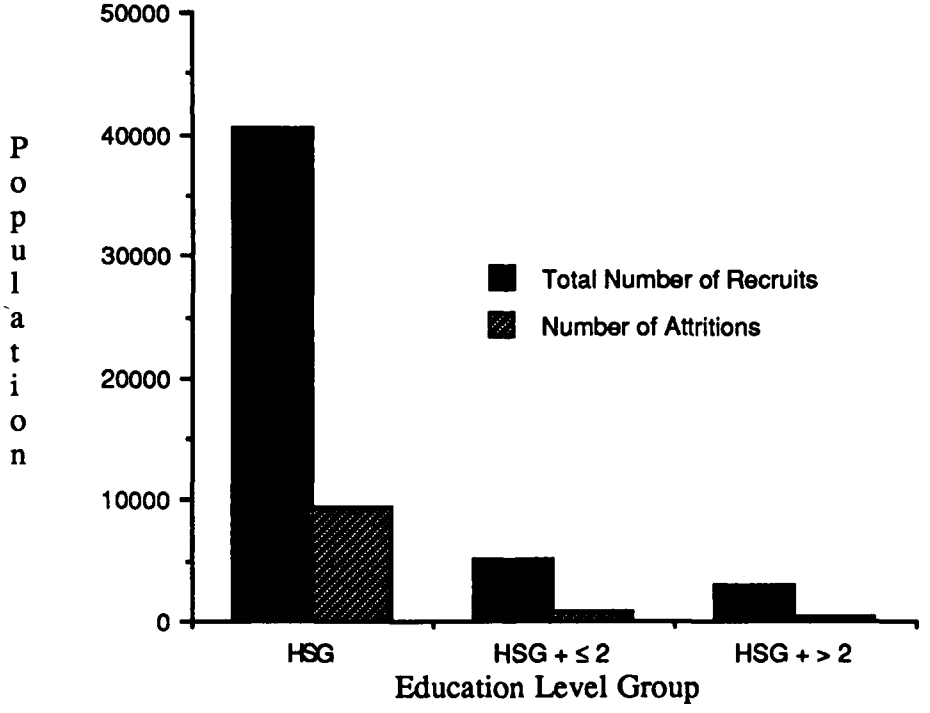

Figure 3.13 : FY84 NPS Accession Cohort - Sample Group - by Education Level Group - Total Numbers vs Numbers of Attrition -

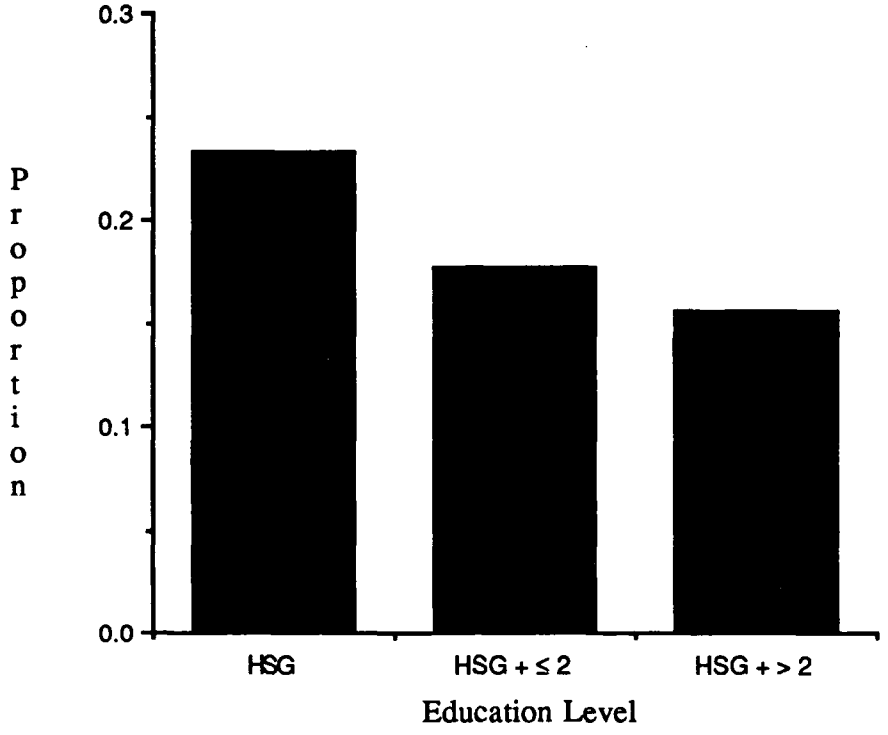

Figure 3.14 : FY84 NPS Accession Cohort - Sample Group - by Education Level Group - Proportion of Attrition

3.3.3 APTITUDE CATEGORY

High quality recruits' aptitude levels ranged from Category IIIA to Category I. [1] The two aptitude levels with the highest enlistment numbers were Category II, with 26028 (53.6%) and Category IIIA, with 17140 (35.3%). The highest attrition rate belonged to the Category IIIA, at 24.5%. Their attrition rates are 21.8% for Category II, and 17.6% for Category I. The attrition rate decreased as the aptitude level increased.

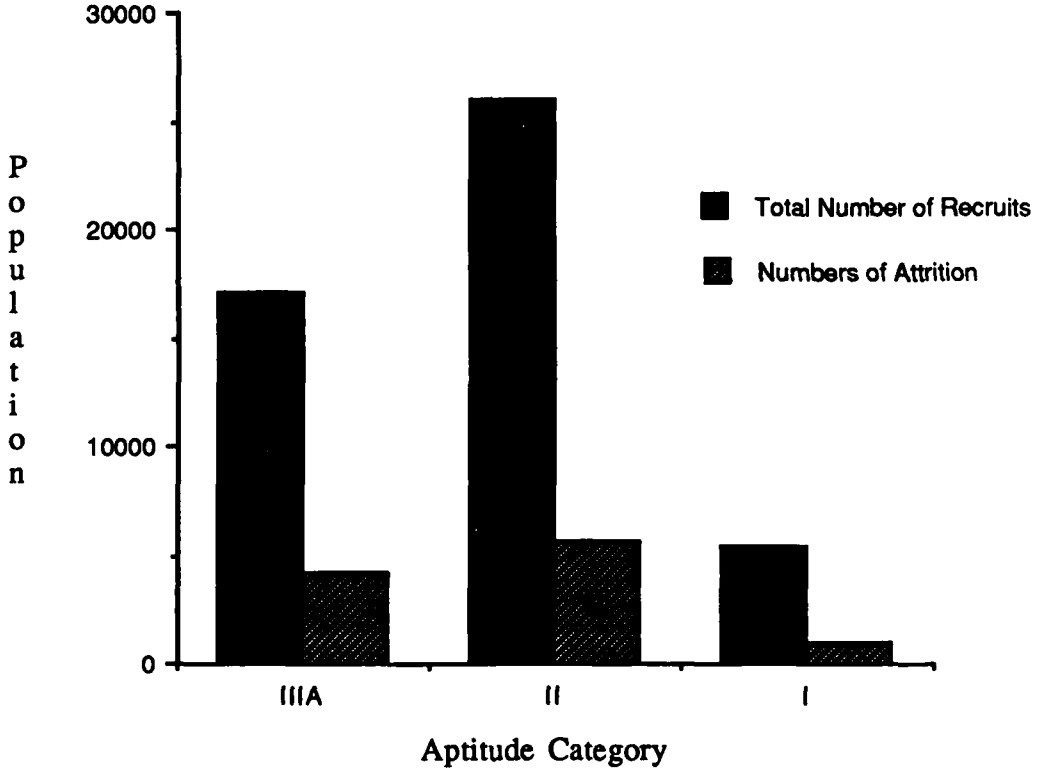

Figure 3.15 : FY84 NPS Accession Cohort - Sample Group - by Aptitude Category - Total Numbers vs Numbers of attrition

---

[1] Cat IIIA : AFQT Score = 50%-64% ; Cat II : AFQT Score = 65%-92%; Cat I : AFQT Score = > 92%

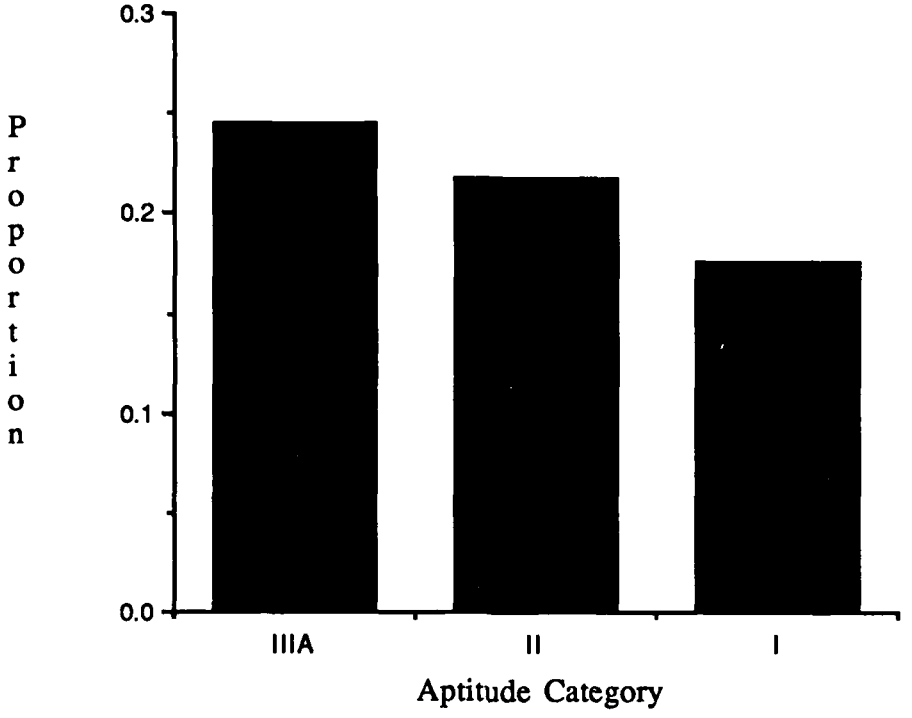

Figure 3.16 : FY84 NPS accession Cohort - Sample Group - by Aptitude Category - Proportion of Attrition

3.3.4 WAIVER STATUS

Recruits came into the Army either with or without a waiver. A person meeting all entry requirements will enter without a waiver. A person desiring to enter the Army who does not meet the normal entry standards, may request a waiver for enlistment. For example, a person not meeting the age limit of 17 to 35 may request a waiver for age. Another can request a waiver for physical qualification if he or she is not deemed physically fit by entry standards. There are numerous waiver categories, as listed in Appendix A. Only a very small group entered with waivers . They numbered 5,500 (11.3%), with an attrition rate of 27.5%. The attrition rate for those who entered the Army without a waiver was 21.6%.

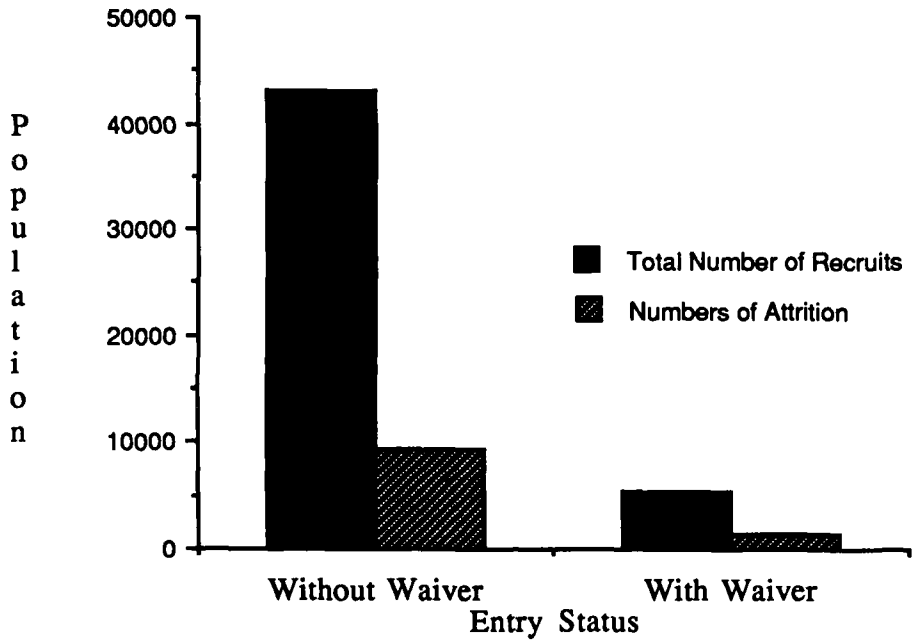

Figure 3.17 : FY84 NPS Accession Cohort - Sample Group - by Entry Status - Total Numbers vs Numbers of Attrition

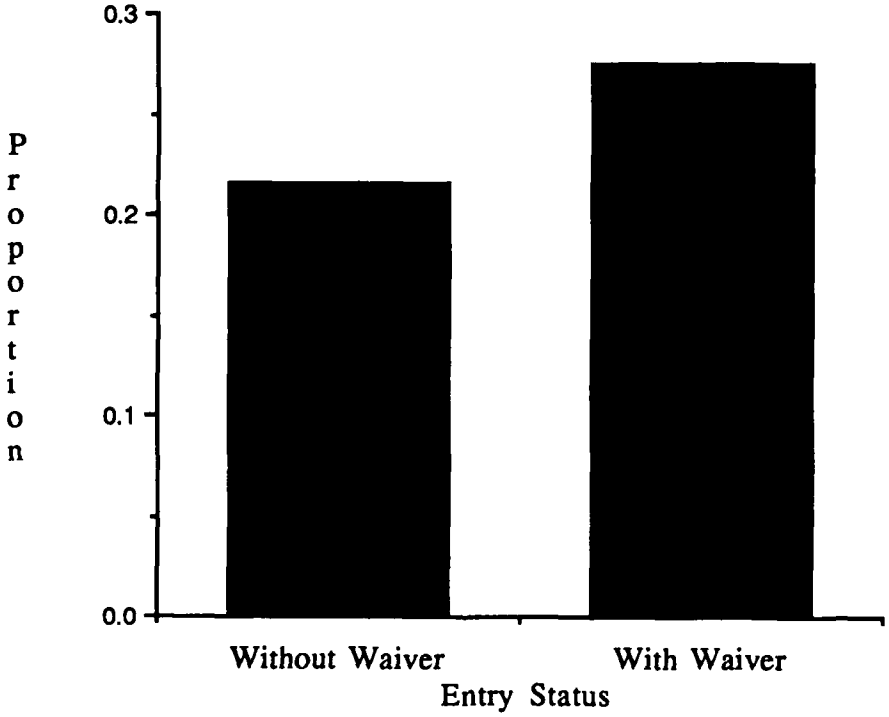

Figure 3.18 : FY84 NPS Accession Cohort - Sample Group - by Entry Status - Proportion of Attrition -

3.4 MODEL ANALYSIS

3.4.1 CHI-SQUARE TEST RESULTS

The Chi-square test strongly rejected the hypothesis that the attrition rates in each of the separate cells were the same. Each cell is a grouping of individuals belonging to the same age group, education level group, aptitude test score category and entry status. There were 4 age groups (X11 , X12, X13, X14), 3 education level groups (X21, X22, X23), 3 aptitude test score category groups (X31, X32, X33), and 2 entry status groups (X41, X42). As a result, we had 72 separate cells. An example of a cell would be X11X21X31X41, meaning the cell consists of individual of age group 17 - 21, high school graduate, aptitude category IIIA, and no waiver required for entry. The test results suggested that the variation seen among the attrition rates around the overall average was not due to random noise. The Chi-square value for our data was 509.4. The 95 th percentile of the Chi-square distribution with 70 degrees of freedom is 88.4.

The Chi-square test is generally very cruel to large sample sizes, and one wonders whether it is the test being harsh on the data, or there is a major problem with the hypothesis. In this case, there was little room for doubt. The discrepancy was huge. The relevant χ^2 distribution has a mean of 70, variance of 140, and standard deviation of 11.83. Thus, the test Chi-square value was over 35 standard deviations above the mean. The test results encouraged further efforts to seek a prediction rule that yields different forecasts for different groups of recruits.

3.4.2 LOGIT AND MLE RESULTS

The model with age, education, aptitude score category and entry status was formulated with the logistic function. MLE was used to obtain the parameter values. The regression model was entered into SYSTAT, as follows:

$$\text{model failure} = \exp(b_0 + b_1*\text{Age} + b_2*\text{Ed Level} + b_3*\text{Aptitude Cat} + b_4*\text{Entry Status}) + (1 + \exp(b_0 + b_1*\text{Age} + b_2*\text{Ed Level} + b_3*\text{Aptitude Cat} + b_4*\text{Entry Status}) \quad (3.1)$$

$$\text{loss} = -\text{count}*(\text{failure}*\log(\text{estimate}) + (1-\text{failure})*\log(1-\text{estimate})) \quad (3.2)$$

where:

failure = 0 if counted as attrition
= 1 if not counted as attrition

b_n = parameter values for n = 0,1,2,3,4

Age = 1 if X11 , 2 if X12, 3 if X13, 4 if X14

Ed Level = 1 if X21, 2 if X22, 3 if X23

Aptitude Cat = 1 if X31, 2 if X32, 3 if X33

Entry Status = 1 if X41, 2 if X42

loss = MLE loss function

count = number of individuals within that cell, associated with failure = 1 or = 0

estimate = value of model failure computed in the prior iteration

Results of the regression are shown at Table 2.

TABLE 2

PARAMETER VALUES FROM REGRESSION

Parameter	Estimate	Standard Error
b_0	-1.190571	0.075746
b_1	0.310982	0.013333
b_2	-0.414799	0.036670
b_3	-0.156133	0.039601
b_4	0.272056	0.037916

The parameter values implied that age and entry status were positively associated with the dependent variable. It meant that higher age and/or entry with a waiver would result in a higher attrition rate. Conversely, education level and aptitude category were negatively associated with the dependent variable. Thus, the higher one's education level and/or aptitude category, the lower one's probability of attrition.

The parameter values were entered into the model's logistic function and all predicted attrition rates were calculated. These rates and their corresponding observed rates are shown in Table 3. There were 45 cells (66.2%) that had predicted values within one standard deviation of their means, 16 cells (23.5%) that had predicted values between one and two standard deviations away, and the remaining 7 cells (10.3%) had predicted values beyond two standard deviations away. This looked good, since with 68 cells, even if the probabilistic prediction was right, about 68% should be within one standard deviation of the mean and 95% within two standard deviations of the mean.

TABLE 3

EXPECTED VS OBSERVED RATES - SAMPLE GROUP

Age = 1 if X11 , 2 if X12, 3 if X13, 4 if X14

Education = 1 if X21, 2 if X22, 3 if X23

Aptitude Category = 1 if X31, 2 if X32, 3 if X33

Entry Status = 1 if X41, 2 if X42

Age	Education	Aptitude Category	Entry Status	Population	Expected Attrition	Observed Rate
1	1	1	1	12843	.235	.233
1	1	1	2	1350	.288	.316
1	1	2	1	16810	.208	.212
1	1	2	2	1740	.257	.272
1	1	3	1	2182	.184	.182
1	1	3	2	240	.228	.167
1	2	1	1	428	.169	.143
1	2	1	2	42	.211	.167
1	2	2	1	1413	.148	.148
1	2	2	2	195	.186	.169
1	2	3	1	580	.129	.129
1	2	3	2	65	.163	.138
1	3	1	1	14	.118	.071
1	3	1	2	1	.150	.000
1	3	2	1	68	.103	.162
1	3	2	2	5	.131	.000
1	3	3	1	47	.089	.064
1	3	3	2	2	.114	.000
2	1	1	1	1047	.296	.289
2	1	1	2	276	.355	.304
2	1	2	1	1757	.264	.272
2	1	2	2	493	.320	.284
2	1	3	1	347	.235	.196
2	1	3	2	96	.287	.281
2	2	1	1	266	.217	.233
2	2	1	2	46	.267	.283
2	2	2	1	832	.192	.181
2	2	2	2	168	.238	.268
2	2	3	1	349	.169	.158
2	2	3	2	86	.210	.291
2	3	1	1	178	.155	.124
2	3	1	2	20	.194	.150
2	3	2	1	828	.135	.118
2	3	2	2	84	.171	.060
2	3	3	1	650	.118	.117

TABLE 3 (Continued)

Age	Education	Aptitude Category	Entry Status	Population	Expected Attrition	Observed Rate
2	3	3	2	67	.150	.134
3	1	1	1	226	.364	.354
3	1	1	2	63	.429	.476
3	1	2	1	444	.329	.315
3	1	2	2	152	.392	.362
3	1	3	1	97	.296	.309
3	1	3	2	29	.355	.241
3	2	1	1	88	.275	.227
3	2	1	2	12	.332	.417
3	2	2	1	224	.245	.246
3	2	2	2	42	.298	.095
3	2	3	1	85	.217	.188
3	2	3	2	19	.267	.263
3	3	1	1	69	.200	.246
3	3	1	2	8	.247	.250
3	3	2	1	317	.176	.192
3	3	2	2	41	.219	.195
3	3	3	1	245	.155	.184
3	3	3	2	32	.194	.219
4	1	1	1	84	.439	.429
4	1	1	2	16	.507	.500
4	1	2	1	129	.401	.388
4	1	2	2	38	.468	.500
4	1	3	1	43	.364	.419
4	1	3	2	3	.429	.000
4	2	1	1	25	.341	.400
4	2	1	2	6	.404	.333
4	2	2	1	75	.307	.293
4	2	2	2	15	.367	.533
4	2	3	1	36	.274	.278
4	2	3	2	9	.332	.111
4	3	1	1	29	.254	.448
4	3	1	2	3	.309	.557
4	3	2	1	142	.226	.338
4	3	2	2	16	.277	.312
4	3	3	1	86	.200	.256
4	3	3	2	20	.247	.300

The residual plot is displayed at Figure 3.19. The distribution of the residuals were fairly random, with a slightly larger spread at the higher cell groups.

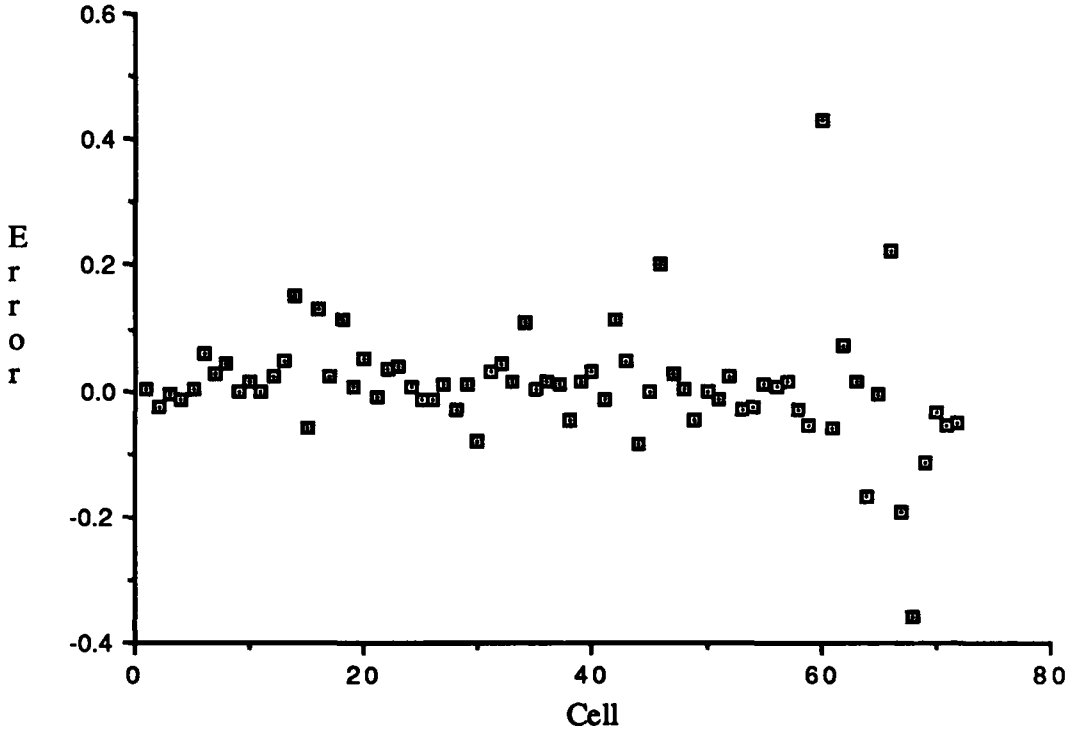

Figure 3.19 : Residual Plot of Expected Attrition Rates
vs Observed Attrition Rates

The model fitted the data very well. The explanatory variables associated positively or negatively, as one would have anticipated, with the dependent variable. A look at the extreme ends of the cells showed that the model behaved excellently there. The highest risk group, per the model, is cell $X_{14}X_{21}X_{31}X_{42}$ (age $\geq$ 30, HS graduate, Cat IIIa, with waiver). It had a predicted attrition rate of 50.7%, versus a 50.0% for the observed rate. The lowest risk group is cell $X_{11}X_{23}X_{33}X_{41}$ (age 17 -21, HS graduate with > 2 yrs of college, cat I, without waiver). It had a predicted attrition

rate of 8.9%, while the observed rate was 6.4%. The differences are not statistically significant.

3.5 MODEL VALIDATION

To validate the original model, the 5% hold-out sample was used. The question is whether the pattern that prevailed in 95% of the whole population will also prevail in the other 5% that had nothing to do with model formulation. If there exists a good fit, then one's confidence in the model increases. It means that the model did more than just " coddle" the data of the sample group.

Only 59 of the 72 cells were used with the hold-out sample. The deleted cells had zero entries. The model also behaved nicely with this set of data. The highest risk group, among the 59 cells, is cell $X_{14}X_{21}X_{32}X_{42}$ (age $\geq$ 30, HS graduate, Cat II, with waiver). It had a predicted attrition rate of 46.8%, versus a 60.0% for the observed rate. The lowest risk group is cell $X_{11}X_{23}X_{33}X_{41}$ (age 17 -21, HS graduate with > 2 yrs of college, cat I, without waiver). It had a predicted attrition rate of 8.9%, while the observed rate was 0.0%. The differences again are not statistically significant. The predicted and observed values are listed at Table 4.

To further examine the accuracy of the model, a goodness-of-fit test between the observed and expected frequencies was conducted. First, we calculated where the observed attrition rates fell as percentiles of their distributions given that the binomial parameters within each group were correct. We did this by using the technique of normal approximation to the binomial distribution (Walpole, 1978) and the z table.

The z values were calculated with :

$$Z = \frac{X - np}{\sqrt{npq}} \qquad (3.3)$$

where:
X = observed number of attrition for the cell
n = total number of recruit for the cell
p = expected probability of attrition for the cell
q = 1 - p
$\sqrt{npq} = \sigma$ = standard deviation for the cell

To ensure that the approximation is fairly good, we combined some cells together in order for $n \geq 4$ per cell. As such, the number of cells was reduced to 28 from 59. We then hypothesized that the distribution of these probabilities is uniform, with 2.8 cells expected in each of the 10 equally spaced percentile cells. The observed frequencies and the expected frequencies per cell are listed in Table 5. Figure 3.20 provides a graphical display of the data.

TABLE 4

EXPECTED VS OBSERVED RATES - HOLD- OUT SAMPLE

Age = 1 if X11 , 2 if X12, 3 if X13, 4 if X14

Education = 1 if X21, 2 if X22, 3 if X23

Aptitude Category = 1 if X31, 2 if X32, 3 if X33

Entry Status = 1 if X41, 2 if X42

" - " = no entries in that cell

Age	Education	Aptitude Category	Entry Status	Population	Expected Attrition Rate	Observed Attrition Rate
1	1	1	1	649	.235	.234
1	1	1	2	64	.288	.297
1	1	2	1	861	.208	.182
1	1	2	2	94	.257	.277
1	1	3	1	127	.184	.197
1	1	3	2	7	.228	.143
1	2	1	1	24	.169	.169
1	2	1	2	3	.211	.333
1	2	2	1	73	.148	.164
1	2	2	2	10	.186	.100
1	2	3	1	38	.129	.000
1	2	3	2	1	.163	.000
1	3	1	1	3	.118	.000
1	3	1	2	-	-	-
1	3	2	1	2	.103	.000
1	3	2	2	-	-	-
1	3	3	1	2	.089	.000
1	3	3	2	-	-	-
2	1	1	1	50	.296	.240
2	1	1	2	16	.355	.312
2	1	2	1	99	.264	.263
2	1	2	2	29	.320	.379
2	1	3	1	23	.235	.217
2	1	3	2	2	.287	.500
2	2	1	1	12	.217	.083
2	2	1	2	3	.267	.333
2	2	2	1	41	.192	.244
2	2	2	2	12	.238	.000
2	2	3	1	20	.169	.050
2	2	3	2	4	.210	.000
2	3	1	1	8	.155	.000
2	3	1	2	1	.194	.000

TABLE 4 (Continued)

Age	Education	Aptitude Category	Entry Status	Population	Expected Attrition	Observed Rate
2	3	2	1	62	.135	.113
2	3	2	2	8	.171	.250
2	3	3	1	32	.118	.094
2	3	3	2	3	.150	.000
3	1	1	1	14	.364	.357
3	1	1	2	-	-	-
3	1	2	1	26	.329	.308
3	1	2	2	9	.392	.556
3	1	3	1	3	.296	.000
3	1	3	2	3	.355	.333
3	2	1	1	8	.275	.375
3	2	1	2	1	.332	.000
3	2	2	1	11	.245	.182
3	2	2	2	3	.298	.000
3	2	3	1	6	.217	.000
3	2	3	2	4	.267	.050
3	3	1	1	3	.200	.000
3	3	1	2	1	.247	1.000
3	3	2	1	20	.176	.150
3	3	2	2	3	.219	.000
3	3	3	1	21	.155	.095
3	3	3	2	-	-	-
4	1	1	1	7	.439	.429
4	1	1	2	-	-	-
4	1	2	1	11	.401	.364
4	1	2	2	5	.468	.600
4	1	3	1	-	-	-
4	1	3	2	-	-	-
4	2	1	1	1	.341	1.000
4	2	1	2	-	-	-
4	2	2	1	2	.307	.000
4	2	2	2	1	.367	1.000
4	2	3	1	2	.274	.000
4	2	3	2	-	-	-
4	3	1	1	-	-	-
4	3	1	2	-	-	-
4	3	2	1	2	.226	.500
4	3	2	2	1	.277	.000
4	3	3	1	6	.200	.333
4	3	3	2	-	-	-

TABLE 5

EXPECTED VS OBSERVED FREQUENCIES - HOLD-OUT SAMPLE

Cell	Percentile	Expected Frequency	Observed Frequency
1	0 - 10	2.8	3
2	10 - 20	2.8	4
3	20 - 30	2.8	5
4	30 - 40	2.8	2
5	40 - 50	2.8	6
6	50 - 60	2.8	1
7	60 - 70	2.8	3
8	70 - 80	2.8	2
9	80 - 90	2.8	2
10	90 - 100	2.8	0

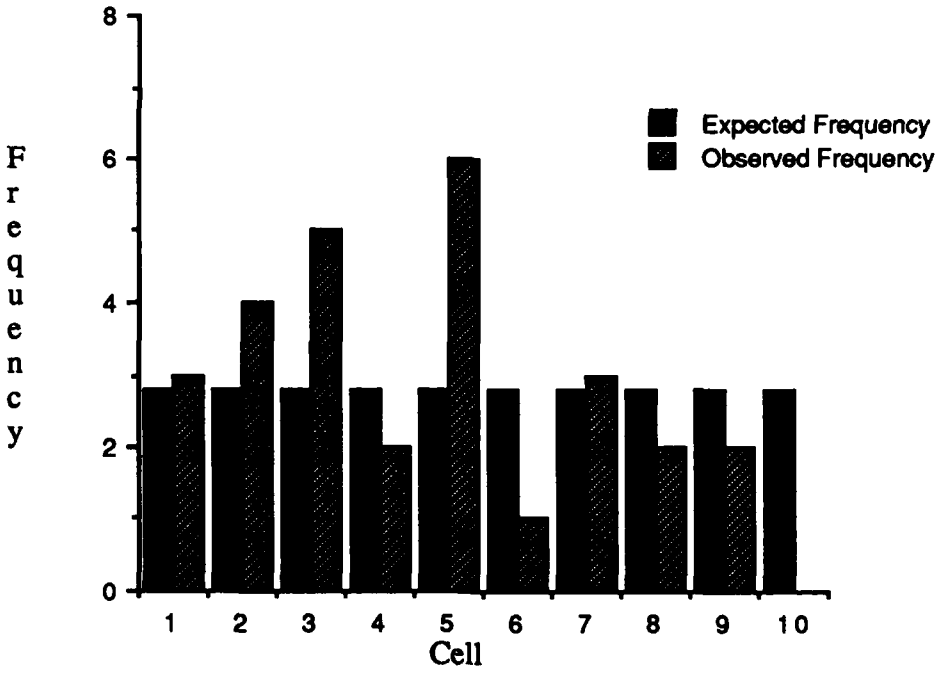

Figure 3.20 : Expected vs Observed Frequencies

The goodness-of-fit test between the observed and expected frequencies is based on the quantity:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad (3.4)$$

where χ^2 is a value of a random variable whose sampling distribution is approximated very closely by the chi-square distribution. o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell (Walpole, 1978). If the observed frequencies are close to the expected frequencies, then the χ^2 value is small denoting a good fit. If the differences are large, then the χ^2 value is large and the fit is poor. The χ^2 value of the model is at 10.57 and the value from the χ^2 distribution is 16.92, at .05 level of confidence, attesting to the accuracy of the model. As seen from Figure 3.20, generally the observed frequency was more than expected at the middle and at the low percentile range. This is consistent with the notion that the model has behaved well, since it meant that more often than not when the model erred, it erred very little.

Chapter 4

CONCLUSIONS

4.1 GENERAL

This logistic regression model did very well in quantifying the effects of the variables considered on attrition levels. It not only provided a good approximation with respect to the sample group, but also behaved well with the 5% hold-out sample, which did not take an active role in the model calibration. However, it must be pointed out that even though this model behaved well, it is only one good model among a possible larger group of good models.

4.2 FACTORS CONTRIBUTING TO ATTRITION

The results of the analysis showed that four significant characteristics related to attrition are: age, level of education, aptitude test score, and entry status.
Age and entry status displayed positive effects on the rate of attrition, i.e., as age increased and/or entry status included a waiver, the attrition risk for that recruit increased. Education and aptitude test score , on the other hand, had negative effects on the attrition rate. The higher one's level of education and/or aptitude test score, the higher the probability that one will remain in service until completion of one's obligated tour. Moreover, a probabilistic model based solely on these four factors fit the data set down to a small level of discrepancy that can be readily explained as statistical noise.

4.3 POLICY IMPLICATIONS AND FUTURE RESEARCH

4.3.1 POLICY IMPLICATION

U.S. Army leaders and manpower planners might find this study to be useful in improving the attrition rate of future NPS high-quality male recruits of an accession cohort. There are several courses of action (CA):

1. Increase enlistment of individuals with the traits of groupings which had lower attrition rates than the cohort attrition rate.

2. Decrease enlistment of individuals with the traits of groupings which had higher attrition rates than the cohort attrition rate.

3. Combine both of the above courses of action.

CA 1 will increase the number of recruits with low attrition rates, thereby decreasing the overall attrition rate of the entire cohort. CA 2 will decrease the number of recruits with high attrition rates, again the net effect is a lower overall cohort attrition rate. CA 3 will have similar effects.

The use of incentives and the increase of emphasis in recruiting a particular segment of the eligible population are legitimate possibilities. The types and scope of the incentives are beyond the objective of this study.

4.2.2 FUTURE RESEARCH

This study looked only at the high-quality male population. The logistic regression model formulation and analysis should be done for the entire population of a NPS

accession cohort. The result, I propose, would be similar, in that age and entry status would display positive associations with the rate of attrition, while education and aptitude test score conversely would have negative effects. This might suggest a recruiting strategy under which readiness posture of the Army is enhanced.

A study should be conducted to determine the most cost effective means to increase the numbers of desired recruits, recommended by this study.

APPENDIX A

WAIVER CATEGORIES

1. Age
2. Number of Dependents
3. Mental Qualification
4. Moral Qualification
5. Previous Disqualification / Separation
6. Lost Time
7. Physical Qualification
8. Sole Surviving Member
9. Education
10. Alien
11. Security Risk
12. Conscientious Objector
13. Pay Grade
14. Skill Requirements
15. Predictor Requirements

BIBLIOGRAPHY

Antel, John, Hosek, James R., and Chritine E. Peterson, *Military Enlistment and Attrition*, The RAND Corporation Report R-3510-FMP, June 1987.

Ashton, Winifred D., *The Logit Transformation with Special Reference to Its Use in Bioassay*, Charles Griffin and Company Limited, London, 1972.

Black, Matthew, and Thomas Fraker, "First-Term Attrition of High School Graduates in the Military," in Curt Gilroy (ed.), *Army Manpower Economics*, Westview Press, Boulder, Colorado, 1986.

Breiman, Leo, *Statistics: With a View Toward Applications*, Houghton Mifflin, Boston, 1973.

Buddin, Richard, *Analysis of Early Military Attrition Behavior*, The RAND Corporation, R-3609-MIL, July 1984.

Buddin, Richard, *Trends in Attrition of High Quality Military Recruits*, The RAND Corporation, R-3539-FMP, August 1988.

Buddin, Richard, and Christina Witsberger, *Reducing the Air Force Male Enlistment Requiremnt - Effects on Recruiting Projection of the Other Services*, The RAND Corporation , R-3265-AF, March 1985.

Cotterman, Robert F., *Forecasting Enlistment Supply*, The RAND Corporation, R-3252-FMP, July 1986.

Cox, D.R., *Analysis of Binary Data*, Chapman and Hall, London, 1977.

Everitt, B.S., and G. Dunn, *Advanced Methods of Data Exploration and Modelling*, Heinemann Educational Books, London, 1983, pp. 154 - 175.

Fujikoshi, Yasunori, *Selection of Variables in Discriminant Analysis and Canonical Correlation Analysis*, (P.R. Krishnaiah, ed.), Multivariate Analysis - VI, North Holland Amsterdam, Proceedings of the Sixth International Symposium on Multivariate Analysis, pp. 219-236.

Grissmer, David W., and Sheila Nataraj Kirby, *Attrition of Non-Prior Service Reservists in the Army National Guard and Army Reserve*, The RAND Corporation, R-3267-RA, October 1985.

Halperin, M., Blackwelder, W.C., and J.I. Vorter, " Estimation of the Multivariate Logistic Risk Function: A Comparison of the Discriminant Function and Maximum Likelihood Approaches," *Journal of Chronic Diseases*, Vol. 24, 1971, pp. 125 - 158.

Hanushek, Eric ., and John E. Jackson, *Statistical Methods for Social Scientists*, Academic Press, New York, 1977, pp. 325 - 344.

Horne, David K., *The Impact of Soldier Quality on Performance in the Army*, Technical Report 708, Manpower and Personnel Research Laboratory, Research Institute for the Behavioral and Social Sciences, April 1986.

Hosek, James R., Peterson, Christine E., and Rick A. Eden, *Educational Expectations and Enlistment Decisions*, The RAND Corporation, R-3350-FMP, March 1986.

May, Laurie J., and Jacquelyn Hughes, *Estimating the Cost of Attrition of First Term Enlistees in the Marine Corps*, Marine Corps Operations Analysis Group CRM86-168, June 1986.

Morris, Carl N., and John E. Rolph, *Introduction to Data Analysis and Statistical Inference*, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

Neter, John, Wasserman, William, and Michael H. Kutner, *Applied Linear Regression Models*, Irwin, Homewood, Illinois, 1983, pp. 328 - 376.

Polich, J. Michael, Dertouzoz, James N., and S. James Press, *The Enlistment Bonus Experiment*, The RAND Corporation, R-3353-FMP, April 1986.

Quester, Aline, and Martha S. Murray, *Attrition from Navy Enlistment Contracts*, Naval Planning Program and Logistic Division CRM86-12, January 1986.

Stolzenberg, Ross M., and John D. Winkler, *Voluntary Terminations from Military Service*, The RAND Corporation, R-3211-MIL, May 1983.

Toomepuu, Juri, *Costs and Benefits of Quality Soldiers*, USAREC Research Note 86-1, September 1986.

Toomepuu, Juri, *Education and Military Manpower Requirements*, Presentation to Annual Meeting of the Council of Chief State School Officers, U.S. Army Recruiting Command, November 15, 1986.

Vinod, Hrishikesh D., and Aman Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, Inc., New York, 1981, pp. 300-302.

Walpole, Ronald E., and Raymond H. Myers, *Probability and Statistics for Engineers and Scientists*, Macmillan, New York, 1978, pp. 189 - 280.

Wilkinson, Leland, *SYGRAPH*, SYSTAT Inc., Evanston, Illinois, 1988.

Wilkinson, Leland, *SYSTAT*, SYSTAT Inc., Evanston, Illinois, 1988.